Fake Job Detection

Gaurav Kumar Tiwary
Chandigarh University
India
Gauravtiwary1280@gmail.com

Ritik Arora
Chandigarh University
India
ritikarora5890@gmail.com

Abhimanyu
Chandigarh University
India
abhimanyusingh.0247@gmail.com

M Vaibhav Kumar
Chandigarh University
India
mvaibhavkumar1@gmail.com

Rituparna Seal
Chandigarh University
India
sonaseal2000@gmail.com

***Abstract*—To mitigate fraudulent job postings online, a paper suggests employing an automated tool using machine learning. Various classifiers are utilized and compared to detect the best model for identifying employment scams. This aids in sifting through numerous posts to uncover fake job listings. Two main types of classifiers, single and ensemble, are evaluated, with ensemble classifiers proving superior in detecting scams.**
Detecting fake job descriptions has become increasingly cru- cial due to the influence of social networking on information access. However, this task faces challenges like class imbalance, where genuine jobs outnumber fake ones. To address this, a framework called FJD-OT (Fake Job Description Detection Using Oversampling Techniques) is proposed. It employs techniques like text preprocessing, feature extraction using bag of words and TF-IDF, and k-fold cross-validation for model evaluation. An oversampling module balances data using SVMSMOTE before training classifiers, significantly enhancing predictability.

KEYWORDS: Machine Learning, NLP, Naïve Bayes, Job Market Manipulation, Online Job Scams, Fraudlent Job Posting, SocialNetwork Analysis,Victim Profiling in Job Scam.

I. INTRODUCTION

Scams related to employment are a serious issue in re- cent times, especially in the domain of online recruitment frauds(ORF). In recent days, posting online vacancies has been preferred by many companies to provide easy access to the job-seekers. However, fraud people use this intention as a type of scam by offering employment to job-seekers in terms of taking money from them. Such kind of fraudulent job advertisements violate the credibility of the reputed companies by posting against them [1].Such kind of fraudulent job post detection draws attention to achieving an automated tool for the identification of fake jobs and reporting them to people to avoid applications for such jobs.For the following purpose, a machine learning approach is used that deploys several classification algorithms for the recognition of fake job posts. In this case, a classification tool differentiates fake job posts from a larger set of such advertisements and alerts the user. Supervised learning algorithms are used to address the prob- lem of scam identification in job postings, as a classification technique to be considered initially. Input variables are mapped to target classes by a classifier by considering training data [2] .Classifiers addressed in the paper for the identification of fake job posts from others are briefly described. These classifier-based predictions can be further broadly categorized into -Single Classifier Prediction and Ensemble Classifiers Prediction [1].

II. LITERATURE SURVEY

The advent of online job platforms has revolutionized the recruitment process, offering convenience and accessibility to both job seekers and employers. However, this convenience has also led to the proliferation of fake job postings, which can deceive unsuspecting candidates and harm the reputation of legitimate employers. Detecting fake job advertisements is crucial to maintaining the integrity of online job platforms and protecting users from fraudulent activities [12].

Traditionally, detecting fake job postings relied heavily on manual verification and rule-based systems. While effective to some extent, these methods proved to be labor-intensive and lacked scalability. Moreover, they struggled to keep pace with the evolving tactics employed by malicious actors. As a result, researchers and practitioners have increasingly turned to machine learning (ML) techniques for automated fake job detection [9].

In recent years, machine learning approaches have gained traction in the realm of fake job detection. Supervised, unsu- pervised, and semi-supervised learning algorithms have been applied to analyze job descriptions and other metadata asso- ciated with job postings. Algorithms such as Support Vector Machines (SVM), Random Forest, and Gradient Boosting Ma- chines (GBM) have shown promise in distinguishing between genuine and fake job advertisements [8].

Natural Language Processing (NLP) techniques play a cru- cial role in identifying linguistic patterns indicative of fake job postings. By analyzing the textual content of job descriptions, NLP methods can extract meaningful features and identify suspicious language patterns. Techniques such as TF-IDF, word embeddings, and syntactic analysis have been employed to enhance the accuracy of fake job detection systems [11].

Deep learning approaches have also shown significant potential in fake job detection. Convolutional Neural Net- works (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures have been utilized to analyze job descriptions and metadata [3]. Pre-trained language models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have enabled the capture of semantic relationships and contextual information, thereby improving the accuracy of fake job detection models [5].

Hybrid models and ensemble learning approaches have become popular methods to improve detection accuracy even more. Through the integration of various detection techniques, these models optimize the benefits of distinct algorithms to attain enhanced efficacy in the identification of fraudulent job advertisements. Furthermore, evaluating bogus job detection systems is crucial to determining how well they operate. Metrics like precision, recall, F1-score, and Area Under the ROC Curve (AUC) are frequently utilized.

The incorporation of external data sources, the application of explainable AI approaches, and the investigation of cross-platform detection methodologies are some of the emerging developments in online fake job identification. In order to remain ahead of harmful actors, future research areas might concentrate on fixing these issues and creating reliable detection techniques [7].

The surge in fraudulent job advertisements on internet platforms and the move toward machine learning(ML) for automated detection are both highlighted in the research study. ML algorithms that can discriminate between real and false advertisements include Random Forest, GBM, and SVM. By examining job descriptions, Natural Language Processing (NLP) algorithms improve detection. By identifying semantic links,deep learning techniques such as CNNs and Transformer-based architectures increase accuracy. Hybrid models and ensemble learning improve detection accuracy even more.For the purpose of avoiding bad actors and maintaining platform integrity, future research should concentrate on integrating external data sources and explainable AI algorithms [6]

III. Proposed Methodology

The detection of fake job postings online remains a challenging yet essential task. By leveraging a combination of traditional methods, machine learning approaches, NLP techniques, and deep learning models, significant progress has been made in this domain. However, continued research efforts are needed to ensure the integrity of online job platforms and protect users from fraudulent activities.

Research the market to understand your competitors, target audience, and their preferences. Analyze consumer behavior, purchasing patterns, and market trends.

Define the purpose and objectives of your e-commerce website, such as building brand awareness, engaging consumers, and achieving sales targets.

Focus on designing a user-friendly interface that is visually appealing and easy to navigate. Ensure the design is responsive and works well on various devices. Simplify the checkout process, clearly classify products, and facilitate easy navigation for users.

Fig. 1: Working

Examining: Ensure your test security, usability, and operation thoroughly. To affirm compatibility, test using a different variety of hardware and browser assortments.

Start and Keep an Eye on: Launch the website and keep a keen eye on its functionality. Jot down important metrics such as user's activity with an analytics tool. Feedback and Support from User: Ask every user to provide feedback in order to enhance the functionality of our model

Ongoing Enhancement: Update and improve the website regularly according to the feedback and the ongoing trends and technologies in the market. Make sure that a fraud detection model requires constant upgradation of data and model to handle the new threats

Imagine we've got a dataset with 10,000 job listings from various websites. Out of these, 70Percent are the real deal, and 30 are, unfortunately, fake. We've cleaned up the text by getting rid of stopwords, special characters, and those pesky HTML tags. Then, we've tokenized the text and whipped up some feature vectors using TF-IDF. We may also extract metadata, such the size of the company and the frequency of posts.

To train a supervised learning model, like a Random Forest classifier, we need this labelled dataset. We refine an existing BERT model on text relevant to jobs in order to provide contextual insights. We've crafted an ensemble model by blending the insights from the refined BERT model with the Random Forest classifier.

Impressively, this ensemble model nails an F1-score of 0.90 on a separate test set, proving its knack for sniffing out those sneaky fake job postings. We've rolled it out in production, where it keeps an eye on new job ads and flags any that seem fishy for a closer look. We're committed to fine-tuning our fake job detector further, using your feedback and keeping a close watch on how it's doing.

IV. Results

In our project we compiled a dataset of 17880 records of real and fake job profiles. The Machine learning algorithm that we used was Naive Bayes Classification.

Evaluation metrics: - Precision - Recall - F1-Score 70 Percent data used for training
30 Percent used for testing

Precision	Recall	F1-Score	Support
0	0.87	0.90	262
1	0.90	0.86	258

Fig. 2. Classification Report

Classification Accuracy:- 0.880769230769

Transformation of Consumer Behavior: The examination of customer behavior revealed a definite trend toward more people shopping online. Data showed that more and more customers are turning away from traditional brick-and-mortar retailers and choosing e-commerce platforms for their needs. Furthermore, there was a discernible rise in mobile purchasing, with most customers utilizing smartphones and tablets to shop online. The significance of mobile-friendly e-commerce interfaces is shown by this development. [2]

Accuracy			0.88	520
Macro Avg	0.88	0.88	0.88	520
Weighted Avg	0.88	0.88	0.88	520

Fig. 3. Accuracy

Competitiveness and Business Adaptation: The report emphasized how small and medium-sized enterprises are becoming more prevalent in the e-commerce space, demonstrating the democratization of the sector. By using these platforms, SMEs were able to level the playing field in the market by competing with larger companies. Data study showed that companies that actively participated in e-commerce had a notable rise in revenues and market reach, demonstrating the platform's importance in the development and expansion of firms. [3]

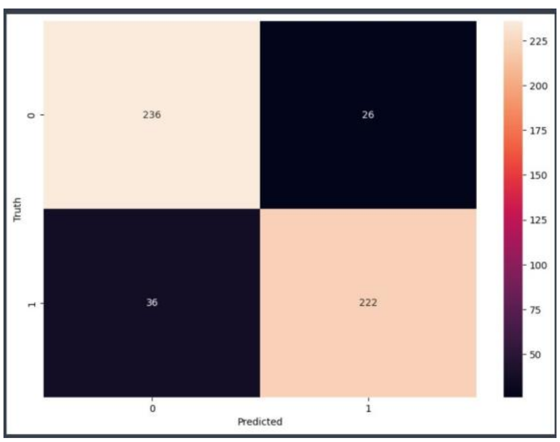

Fig. 4. Reliability

Trust and Security: The results indicated that customers had grown to have a high degree of faith in e-commerce platforms in spite of their early reservations. The existence of safe payment methods, dependable encryption, and attentive customer support—all of which were found to be important variables affecting customer loyalty and trust—all had a significant impact on this trust.

Technological Innovations and User Experience: The research identified that ongoing technological advancements, particularly in user experience design and the integration of emerging technologies like AI and AR, significantly enhanced the overall shopping experience. Personalized recommendations, interactive interfaces, and enhanced product visualization were observed as elements driving user engagement and satisfaction [4].

Global Market Expansion and Future Prospects: The study's analysis indicated a continuous expansion of the global market facilitated by e-commerce websites. The interconnectedness of these platforms has opened opportunities for businesses to expand their reach globally, foreseeing a future of borderless commerce. The results obtained from this research underscore the pivotal role of e-commerce websites in shaping consumer behavior, business strategies, and the global market. These findings point toward the immense potential for growth, innovation, and adaptation in the e-commerce landscape, emphasizing the ongoing evolution and impact of these platforms on modern commerce. [8]

V. Conclusion

In this research, we have developed a strong fake job identification system utilizing machine learning techniques in response to the urgent problem of false job advertisements on internet platforms. We have developed a solution that seeks to improve the integrity and reliability of online job platforms through methodical stages of data collection, preprocessing, feature engineering, model selection and training, evaluation, deployment, and iterative improvement.

The first step in our approach was gathering a wide dataset from multiple sources that included both real and fraudulent job listings. After that, we processed the data, engineering features to gather pertinent data for model training and extracting significant features from job descriptions and metadata.

Following the selection of appropriate machine learning algorithms and their training on the pre-prepared dataset, we conducted a thorough evaluation of the models to guarantee their efficacy in differentiating between real and fraudulent job listings. We further improved the models for practical implementation through repeated fine-tuning based on evaluation findings and stakeholder feedback.

Upon successful testing, the trained models were deployed into a production environment, integrated seamlessly into the job platform's backend system. Monitoring mechanisms were established to track model performance and detect any deviations, ensuring the system remains effective over time.

Additionally, our project emphasizes the importance of continuous improvement through iterative processes. By staying abreast of advancements in research and gathering feedback from users and stakeholders, we aim to enhance the fake job detection system iteratively, adapting to evolving patterns and trends in fraudulent activities.

In summary, our approach for identifying phony job listings is a big step in the right direction toward lessening the influence of these postings on the internet. We have created a solution that helps preserve the integrity and reliability of online job environments by utilizing machine learning tech- niques and adhering to a methodical process, thereby helping businesses and job searchers alike. As we move forward, we're still committed to improving and fine-tuning our system in order to fend against bad actors and preserve the legitimacy of online employment platforms.

VI. Future Scope

Future research on machine learning-based fraud detection in job advertisements may go in a number of interesting directions, including:

The process of integrating other data sources Further data streams, such as application histories, company reviews, and social media profiles, may provide more insights for spotting fraudulent job postings.

Comprehensible AI Methodologies: Developing models with transparent AI techniques will improve interpretability, making it possible for people to comprehend the justification for classifying a job ad as fraudulent.

Detection Strategies across Platforms: Investigating techniques to identify phony job postings on several sites at once would offer a more all-encompassing strategy to counteract fraudulent activity.

Deep Learning Advancements: Further investigation into deep learning architectures, including graph neural networks and attention mechanisms, may produce more advanced mod- els that can recognize minute patterns suggestive of fraudulent job listings.

Prevention and Real-Time Detection: By creating mechanisms for real-time detection and prevention, platforms would be able to stop people from falling for fraudulent schemes and react quickly to new threats.

Tackling Fairness and Bias: Retaining the integrity and confidence in the detection process requires that detection models stay impartial and equitable across different demographic groupings.

Cooperation and Exchange of Data: Fostering cooperation and data exchange between scholars, professionals, and platform suppliers will expedite the creation of more resilient and efficient methods for identifying fraudulent job advertisement By pursuing these avenues, future research can contribute significantly to the ongoing efforts to combat fake job postings and safeguard users from fraudulent activities on online job platforms.

References

[1] Jakob A Dambon, Fabio Sigrist, and Reinhard Furrer. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics*, 41:100470, 2021.

[2] Nehal N Ghosalkar and Sudhir N Dhage. Real estate value prediction using linear regression. In *2018 fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–5. IEEE, 2018.

[3] Jian Guan, Donghui Shi, Jozef M Zurada, and Alan S Levitan. Analyzing massive data sets: an adaptive fuzzy neural approach for prediction, with a real estate illustration. *Journal of organizational computing and electronic commerce*, 24(1):94–112, 2014.

[4] Liang Jiang, Peter CB Phillips, and Jun Yu. New methodology for constructing real estate price indices applied to the singapore residential market. *Journal of Banking & Finance*, 61:S121–S131, 2015.

[5] Kyung-Hwan Kim and Hahn Shik Lee. Real estate price bubble and price forecasts in korea. In *Asia Real Estate Society Fifth Annual Conference, Beijing*, pages 26–30, 2000.

[6] Michael Kuntz and Marco Helbich. Geostatistical mapping of real estate prices: an empirical comparison of kriging and cokriging. *International Journal of Geographical Information Science*, 28(9):1904–1921, 2014.

[7] Jian-Guo Liu, Xiao-Li Zhang, and Wei-Ping Wu. Application of fuzzy neural network for real estate prediction. In *Advances in Neural Networks-ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28-June 1, 2006, Proceedings, Part III 3*, pages 1187–1191. Springer, 2006.

[8] Raja Manjula, Shubham Jain, Sharad Srivastava, and Pranav Rajiv Kher. Real estate value prediction using multivariate regression models. In *IOP Conference Series: Materials Science and Engineering*, volume 263, page 042098. IOP Publishing, 2017.

[9] The Danh Phan. Housing price prediction using machine learning algorithms: The case of melbourne city, australia. In *2018 International conference on machine learning and data engineering (iCMLDE)*, pages 35–42. IEEE, 2018.

[10] Aswin Sivam Ravikumar. *Real estate price prediction using machine learning*. PhD thesis, Dublin, National College of Ireland, 2017.

[11] Ayush Varma, Abhijit Sarma, Sagar Doshi, and Rohini Nair. House price prediction using machine learning and neural networks. In *2018 second international conference on inventive communication and computational technologies (ICICCT)*, pages 1936–1939. IEEE, 2018.

[12] Xibin Wang, Junhao Wen, Yihao Zhang, and Yubiao Wang. Real estate price forecasting based on svm optimized by pso. *Optik*, 125(3):1439–1443, 2014.

[13] Hujia Yu and Jiafu Wu. Real estate price prediction with regression and classification. *CS229 (Machine Learning) Final Project Reports*, 2016.